

On the Role of Individual Items in Recognition Memory and Metacognition: Challenges for Signal Detection Theory

Thomas A. Busey and Anne Arici
Indiana University, Bloomington

The authors tested the role of individual items in recognition memory using a forced-choice paradigm with face stimuli. They constructed distractor stimuli using morphing procedures that were similar to two parent faces and then compared a studied morph against an unstudied morph that was similar to two studied parents. The similarity of the parent faces was carefully balanced so that the choosing rates for the studied and unstudied morphs were approximately equal. Despite being equally likely to choose the studied and the unstudied morph, participants were systematically more confident when choosing the studied morph. This result is incompatible with Gaussian signal detection theory, even with unequal variances for targets and distractors. The authors propose an extension of an extant sampling model, SimSample, which provides a qualitative and quantitative account of the confidence and recognition dissociation. The results suggest that observers make contact with individual items when making recognition judgments with faces and that the structure of the sampling and decision process naturally leads to this dissociation of confidence and recognition.

Keywords: signal detection theory, recognition memory, faces, metacognition

A foundational assumption of traditional signal detection theory (SDT) is that evidence is accumulated along a single decision axis (Green & Swets, 1966). Researchers who use SDT as a model for the decision processes in memory experiments thus make the tacit argument that while the processes that govern a particular memory query might be quite complex, ultimately the evidence in favor of having seen an item before is reduced to a single dimension and compared against a decision criterion (for old/new recognition) or the evidence obtained from two choices is compared (as in two-alternative forced-choice tasks).

The simplicity of the structure of SDT has led to its wide application, as well as to a spirited debate over whether it must be extended to include separate recall-like processes (e.g. Arndt & Reder, 2002; Cary & Reder, 2003; Curran & Cleary, 2003; Jacoby, 1991; Meissner, Tredoux, Parker, & MacLin, 2005; Mickes, Wixted, & Wais, 2007; Rugg et al., 1998; Wixted, 1992; see Yonelinas, 2002, for review). In several cases, authors have explicitly extended SDT into multiple dimensions (Banks, 2000; Parks & Yonelinas, 2007; Rotello, Macmillian Hicks & Hautus, 2006; Rotello, Macmillan & Reeder, 2004). Despite these extensions, some of which violate the unidimensionality assumption, the standard SDT model has proven surprisingly robust (Berry,

Shanks, & Henson, 2008; DeCarlo, 2007; Wixted, 2007a, 2007b; Wixted & Stretch, 2004), and the combination of shifting decision criteria, mixture distributions, and items with distributions that have unequal variances can account for a wide range of data (see Wixted, 2007a, for review).

Our goal in this article is to provide a direct property test of the unidimensional assumption of SDT, which states that target and distractors have values computed along a single evidence axis. We begin with the suggestion that memory is multidimensional, but in the process of having to make a decision, participants in memory experiments collapse the evidence on multiple dimensions down to a single value along a single decision axis. This dimensionality reduction simplifies the comparison against a decision criterion but leaves open the possibility that there may be multiple ways to arrive at a single point on the decision axis. We will explore this possibility using both recognition and confidence judgments as means to document the nature of the decision process.

Several authors (Backus & Nolt, 2001; Loftus & Ruthruff, 1994) have made the point that two stimuli may be *metameric* such that even though they may differ physically, they produce identical values on some internal evidence axis. Color metamers are the best example of this phenomenon because two lights that have very different distributions of energy at different wavelengths might still produce equal ratios of cone captures for the long, medium, and short wavelength cones (Wyszecki, 1958). In more complex settings, two conditions that are equated on one dimension need not be equivalent on others. A simple metaphor would be two climbers who reach the top of Mount Everest and even though they are equated in terms of altitude, they still remember that one reached the top from the North Ridge and the other from the South Col. Likewise two experimental conditions could produce equivalent performance as a result of different processes, and this may reveal the architecture of memory. Such a finding would certainly

Thomas A. Busey and Anne Arici, Department of Psychology, Indiana University, Bloomington.

This work was supported by National Institute of Mental Health Grant 1R03MH060619-01.

We would like to thank Caren Rotello and Richard Shiffrin for advice and discussions of this work, along with Stephen Lewandowsky, Jason Arndt, and Tim Curran for helpful comments during the review process.

Correspondence concerning this article should be addressed to Thomas A. Busey, Department of Psychology, Indiana University, Bloomington, IN 47405. E-mail: busey@indiana.edu

provide evidence against a unidimensional model as a complete explanation.

It was our aim to produce two stimuli that were equated on the evidence axis and then test whether all other information about how each stimulus arrived at that evidence state was indeed lost. To establish two conditions that might produce equivalent evidence states from different sources of information, consider the stimuli in Figure 1. We used morphing procedures (described later) to create morphs from pairs of “parent” faces. Of the eight morphs we created, four were presented on the study list along with the parent faces from the remaining four morphs. At test, we again presented one of the studied morphs along with one of the unstudied morphs in a forced-choice design.

The studied morph would look familiar because it was presented on the study list. The unstudied morph was of course not studied, but its parents were. Thus even though it was not studied, it would have two somewhat weaker matches to memory, which in some cases may be enough to compensate for the fact that it was not actually on the study list. Both the studied and unstudied morphs would also have weak matches to the rest of the study list; although, because of counterbalancing, this fact can be ignored for now (but not in our modeling).

The fact that the unstudied morph has two nearby matches in memory might provide up to perhaps twice as much evidence that the unstudied morph was presented on the study list, despite the fact that the studied morph was presented and the unstudied was not. This is a classic prototype effect (Franks & Branford, 1971; Neumann, 1977; Solso & McCarthy, 1981). We will not argue for prototype formation in this article because we believe that individual exemplars are sufficient to account for these effects (Zaki & Nosofsky, 2001). The likelihood of choosing the studied or unstudied morph depends on the similarity of the two parent faces (and therefore their similarity to the unstudied morph). We could have, for example, used parent faces that were very dissimilar, leading to very low choosing rates for the unstudied morph.

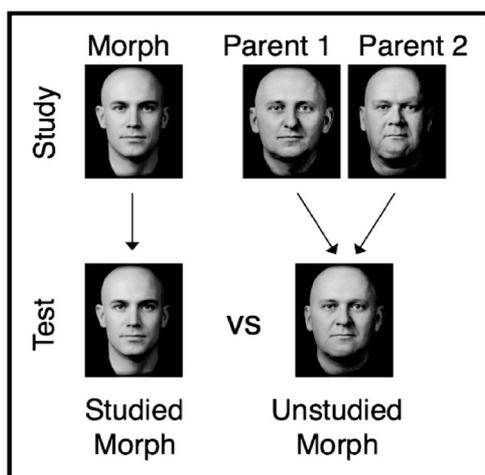


Figure 1. Experimental design and sample stimuli. Eight pairs of parent faces were chosen for their similarity and then used to create morphs. Only two morphs are shown here. Four morphs were studied, while four others had just their parents in the study list. Each critical test trial paired a studied morph with an unstudied morph.

Our forced-choice design has a number of useful properties in that it is criterion free and provides data that we argue are fundamentally incompatible with Gaussian SDT and by implication with all models that reduce recognition decisions to a single dimension. The logic of our design required that we somehow balance the evidence for the studied and unstudied morphs (which could be done by carefully choosing pairs of parent faces that were similar but not too similar), as well as data from confidence ratings, as described next.

Evidence Disambiguation: Confidence Ratings

Evidence that is accumulated for both studied and unstudied morph faces may eventually be reduced to individual values on a single decision axis. However, this does not imply that the details of the individual matches are lost. As with the mountaineering example given earlier, additional queries of the climbers would establish their routes and provide more information beyond the fact that they reached an equivalent altitude. We can acquire the memorial equivalent of route information from confidence ratings.

Generally speaking, a positive correlation between recognition accuracy and confidence is observed and has been described by Hart (1967) as *trace access theory*. However, the literature contains a variety of dissociations between confidence and accuracy. One broad class includes heuristics, whereby participants bring additional information to bear on a memory decision (“Two similar scenes are hard to tell apart,” “Dim faces always produce worse performance at test than bright faces”); Busey, Tunnicliff, Loftus, & Loftus, 2000; Koriat, 1993; Tulving, 1981; see Koriat, 1997, for review), while others rely on different decision criterion for two conditions or different distributional variances for several types of stimuli (Wixted & Stretch, 2004).

In our design, participants may be influenced by the nature of the evidence that underlies support for the studied and unstudied morphs, and this may be reflected in their confidence judgments. For example, they may feel that when they choose the unstudied morph over the studied morph, they are doing so on the basis of an aggregation of lots of weak sources of evidence rather than on the strength of a single strong source. This awareness may not even reside at a conscious level but may be sufficient to produce lower confidence ratings when participants choose the unstudied morph. Such a finding would suggest that the memory process somehow makes contact with individual items in memory, at least to the point at which it would influence a metacognitive judgment. This would be in contrast to global familiarity models—for example, the Generalized Context Model (GCM; Nosofsky, 1986); Search of Associative Memory (SAM; Gillund & Shiffrin, 1984; Raaijmakers & Shiffrin, 1980, 1981), Minerva 2 (Hintzman, 1986, 1988), Composite Holographic Associative Recall Model (CHARM; Eich, 1982, 1985), Theory of Distributed Associative Memory (TODAM; Murdock, 1982), and Retrieving Efficiently From Memory Model (REM; Shiffrin & Steyvers, 1997)—that accumulate evidence across all items and in the process discard information about individual items, at least in some versions of the models.

Balancing Evidence for Studied and Unstudied Morphs

Our goal was to directly test the unidimensional assumption of SDT using a set of conditions in which the evidence for each of

two faces was equated. Four studied morphs and eight parents of the remaining four unstudied morphs were embedded in a larger study list of 14 other faces to give a study list of 26 faces. If our evidence-balancing procedures worked, the choosing rate between the studied and unstudied morphs would be very close to 50%. Note that in traditional forced-choice paradigms, a finding of a 50% choosing rate would imply chance performance and therefore no information about which item was studied. However, as discussed previously, we could take advantage of prototype effects to make the choosing rate vary over a wide range of values simply by adjusting the similarity of the parents to each other (and therefore to the morph). We used stimuli from related experiments (Busey & Tunnicliff, 1999), and so we were able to use the false alarm rates from a larger set of 16 morph/parent sets to select faces that we thought were likely to produce equivalent choosing rates when studied and unstudied morphs were tested together. Whether we succeeded or not is an empirical question.

The critical comparison comes from the studied and unstudied morphs, which due to counterbalancing differed only in whether the morph or its parents were on the study list. We will show that we can produce equal choosing rates for the two types of morphs, yet find that participants gave systematically higher confidence judgments when they choose the studied morph. We also will show how this result is fundamentally incompatible with the assumptions of even unequal-variance SDT and describe an alternative model that can provide a qualitative and quantitative account of this confidence–recognition dissociation.

Method

Stimuli

The stimuli consisted of 60 photographs of bald White men (Kaysler, 1985). All the photographs showed faces at the same frontal view and angle, under the same lighting conditions. All the photographs showed men who were clean shaven and had similar expressions. Sixteen parent faces were used to create eight morphs and were included in the 60-face total list.

The photographic images were displayed on a 21-in. (53.34-cm) Macintosh (Apple, Inc., Cupertino, CA) gray-scale monitor with luminance control and gamma correction from a video attenuator and the VideoToolbox software library (Pelli & Zhang, 1991). The background luminance was set to 5 cd/m². The gray-scale values in the images were scaled to cover the range between 5 cd/m² for black and 80 cd/m² for white. We collected data using one numeric keypad per participant and a single PowerMac computer (Apple, Inc.).

Participants

A total of 502 participants from Indiana University completed this experiment for partial fulfillment of course requirements. Participants were tested in small groups of approximately 4–6 people, with a total of 90 groups. This large number of participants should not be interpreted as an indication that participants were not tested properly or that studies were conducted in a classroom setting. All participants were carefully tested in small groups in the laboratory in experiments that lasted between 30 and 60 min. We chose the large number of participants because each participant

contributed only four trials to the critical comparison, and we were interested in the number of responses that fell in each of six confidence bins. This arrangement tends to spread the data from each participant fairly thinly. In addition, a portion of the logic of our design requires arguing from the null hypothesis, and here a large amount of data is particularly useful. As will be discussed later, we can use statistical resampling procedures to show that our results would have been reliable even had we tested a much more modest set of observers.

Design and Procedure

The study list consisted of 26 total faces that were presented in random order for each group of subjects. Each face was presented for 1,200 ms, followed by a 2-s delay before presentation of the next face. At the start of the study session, participants were told to study each face in preparation for a following memory test. The study list consisted of 4 morphs, 8 parent faces that were used to make the 4 morphs that were not studied, and 14 additional filler faces that were unrelated to the parents or morphs. Fourteen additional filler faces were reserved for use as distractors at testing. The counterbalancing scheme created two study lists such that the morphs were paired together: If a morph was studied for one group and tested against an unstudied morph, the roles were reversed for a second group. This ensured that each pair of morphs spent approximately equal time as a studied and an unstudied morph. The pairings for the filler faces was performed randomly, but again each filler face spent approximately equal amounts of time as a target or a distractor.

The strength of this design is that each morph pair had a choosing rate unique to that pair. This rate may have varied from pair to pair, but we were most interested in the overall choosing rates across all morph pairs.

Immediately after the study phase, participants were tested in a forced-choice recognition paradigm. At test, a studied morph was compared against an unstudied morph in a forced-choice test. There were a total of 18 test trials in the forced-choice recognition test: 4 morph/morph pairs and 14 target/distractor pairs that were presented in random order. Morphs were paired prior to all data collection, and thus the two morphs in a pair alternated as target and distractor. The left–right presentation of studied and unstudied morphs was counterbalanced across participants, as was the location of target filler faces.

In an effort to minimize the possibility that recognition and confidence might be based artificially on different sources of information, we combined the recognition and confidence responses into a single key press on a numeric keypad. At test, the two images of the forced-choice trial were presented on the top of the monitor, and instructions were provided for the single response. Figure 2 illustrates a sample screen. Participants used the 1, 4, and 7 keys on the left side of the numeric keypad to indicate their degree of confidence that the face on the left had been presented on the study list. The 7 key was designated as representing the most confident rating, while 1 key represented the least confident rating. Likewise, the 3, 6, and 9 keys were used to indicate the participants' degree of confidence that the face on the right had appeared on the study trial. Participants were told that there was always a correct answer (a true statement) and that their

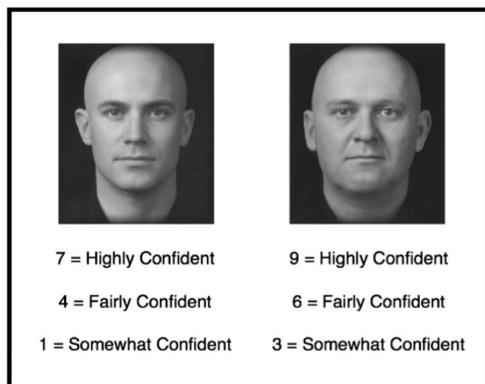


Figure 2. Example forced-choice testing screen in which both recognition and confidence judgments were combined into a single key press. The numbers map to the left and right sides of the numeric keypads used by participants, with higher numbers (both numerically and physically on the keyboard) corresponding to greater degrees of confidence.

single response on each trial indicated both their choice and their degree of confidence that the face appeared.

We anticipated the possibility that if we found differences between studied and unstudied morphs, these differences might be attributable to differences between recollective and familiarity-based processes. Therefore, we asked participants to make recollection/familiarity judgments once all participants had made their combined recognition and confidence judgment. We used an amended recollection/familiarity task, in which we asked participants to press the 5 key if they “recollected the face or a feature” or to press the 2 key if the “face just looked familiar.” Once all participants made a response, the program proceeded to the next test trial. The ordering of all forced-choice test trials was randomized differently for each group of subjects.

Results and Discussion

For the filler faces (those faces that are not studied or unstudied morphs), participants performed at 82.9% accuracy, which corresponds to a d' value of 1.91. This result implies that participants were able to distinguish filler target faces from filler distractor faces. Thus, our participants appeared to understand the task and could perform it at fairly high accuracy with these filler stimuli.

We tested for the possibility that our participants showed a left–right bias but found no evidence for such a bias. For the morphs, the difference between left and right presentations was 0.002, with a 95% confidence interval (–0.043, 0.047) that included zero. Thus, our participants did not seem to favor one side of the display over the other.

Accuracy Comparisons

The critical comparison in this experiment came from test trials in which a studied morph was tested against an unstudied morph. In this crucial test, the probability of choosing the studied morph over the unstudied morph was 0.4994. This was clearly not significantly different from 0.5, and even if it were, it would be in the wrong direction because a number less than 0.5 implies that participants on average chose the unstudied morph over the studied

morph. Thus, in head-to-head testing, observers were equally likely to choose the studied or unstudied morph, which implies that our pilot testing was successful in balancing the similarity of the parent faces such that the unstudied morph seemed about as familiar as the studied morph to the participants. This result should not be interpreted as one in which observers had no memory for the studied morph but instead as one in which the studied and unstudied morphs competed for responses (as will be discussed later). Recall that there was relatively high accuracy for filler faces, which is consistent with participants’ generally understanding the task and performing well on trials that mimicked traditional forced-choice memory experiments.

Despite this equal choosing rate, we observed that observers were more confident when choosing the studied morph than when choosing an unstudied morph, as the following comparisons illustrate. We converted the confidence ratings to a 1–3 scale and found that the average confidence rating when the observers chose a studied morph was 2.13; the average confidence rating when they chose the unstudied morph was 2.02. To assess the reliability of both of these effects in a distribution-free manner, we conducted resampling procedures (Efron & Tibshirani, 1994). We chose this technique as an alternative to traditional null hypothesis testing because it allowed for simulations with fewer numbers of participants (as will be described in a later section) and also allowed us to avoid the rather tortured logic that is required when one attempts to apply null hypothesis testing to a condition in which we expected no differences (and indeed actively worked to eliminate differences), as with the choosing rates. We will show, using histograms, the extent of variability in different conditions, with the intent that the readers should develop their own impression of the data on the basis of inspection of the graphs and confidence intervals that are constructed on the basis of the graphs. Confidence intervals derived from the resampling distributions allow for inferential conclusions (Masson & Loftus, 2003).

The theoretical basis for resampling comes from the assumption that the participants in the experiment are drawn from a larger population and therefore serve as proxies or representatives of this larger group. To establish the reliability of our results, we could resample from our original data with replacement (Efron & Tibshirani, 1994). We resampled the same number of participants as in the original experiment but did so with replacement so that a given participant might be included multiple times or perhaps not at all. Once we had a set of simulated data, we could compute the same summary scores described previously for each simulated experiment. This procedure was repeated 10,000 times, and a histogram was used to illustrate the variability across simulated experiments as well as to construct 95% confidence intervals.

The top-left panel of Figure 3 shows the histograms for the probability of choosing the studied morph, averaged over morphs, and participants for each simulated experiment. As can be seen, the choosing rates centered on .5, and the 95% confidence interval (which we found by identifying the 2.5th and 97.5th percentile values on the sorted scores from all 10,000 resampling experiments) was 0.477 and 0.520. This interval clearly includes 0.5, and thus we are at least 95% confident that the choosing rate for studied morphs fell within a range that represented equality between the studied and unstudied morphs. Of course, our best guess was that the actual number was very close to 0.5.

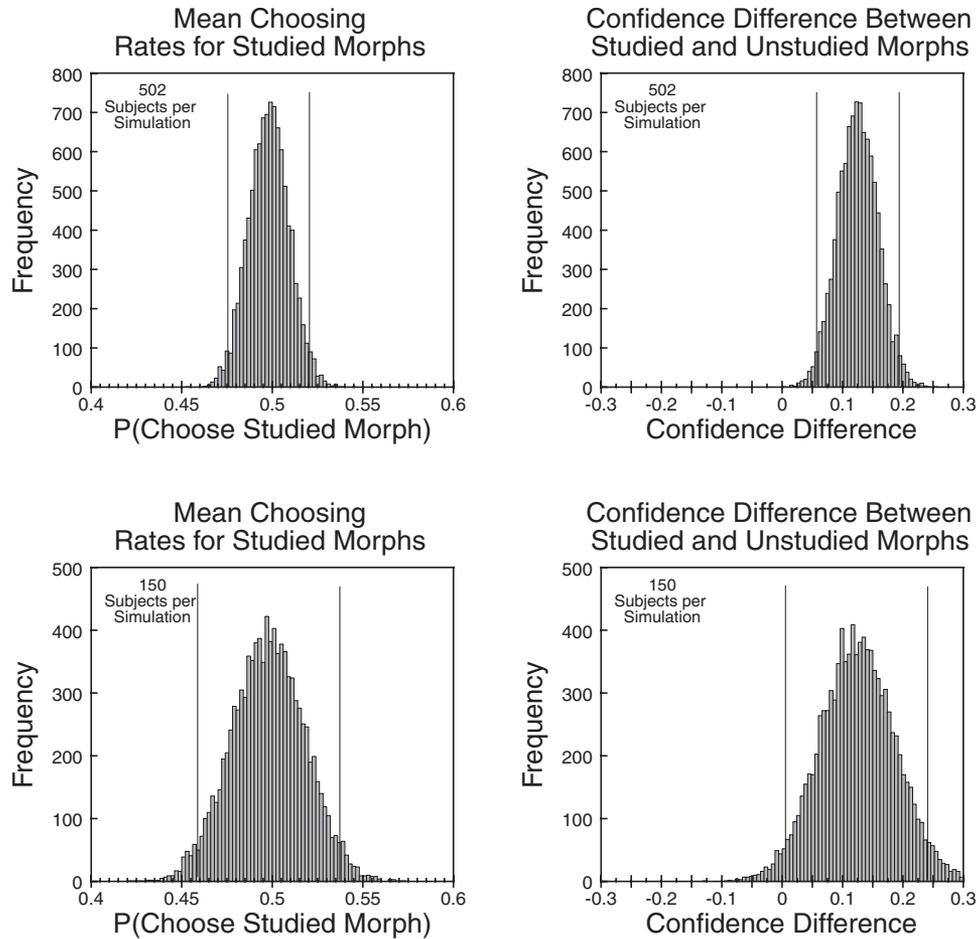


Figure 3. Results of resampling procedures that demonstrate the robustness of the two key findings: the choosing rate is centered on 0.5, and participants are systematically more confident when choosing the studied morph. Top-left panel: Sampled values of choosing rates for the studied morph versus the unstudied morph. Top-right panel: Confidence difference between studied and unstudied morphs. Thin vertical lines are the 95% confidence values for each distribution. Lower panels: Equivalent graphs for simulations with just 150 participants, showing that a large number of participants is not necessary to see the difference in confidence.

The top-right panel of Figure 3 shows the histogram for the difference in confidence between the studied and unstudied morphs when each is chosen. Values to the right were consistent with participants expressing more confidence when they chose the studied morph, and a value of zero would be consistent with no confidence difference when participants choose the studied or unstudied morphs. We observed that the distribution of confidence difference scores in fact shifted in a positive direction, and the 95% confidence interval was 0.060 and 0.191, which did not include zero. In fact, out of the 10,000 simulations, no simulated experiment produced a confidence value that was less than zero. This led us to conclude with 95% confidence that there was a difference in confidence between the studied and unstudied morphs when each was chosen.

One might question whether our results depend on the atypically large number of participants used in our experiment. The resampling procedure allowed us to identify whether we would have obtained a similar conclusion with fewer participants, which we

could do by reducing the size of each simulated experiment from 502 (the number in the actual experiment) to a smaller number such as 150. Because each participant contributed only four trials to the critical comparison, 150 seemed a reasonable number of participants to expect to test in order to obtain reliable results. As shown in the lower panels of Figure 3, resampling fewer participants increased the spread of the simulated means, as one might expect. However, the underlying message was the same: The confidence interval for the choosing rates still included 0.5 (0.458, 0.538) and the 95% confidence interval around the confidence differences still did not include zero (0.002, 0.248).

The results of these simulations are clear: Participants were equally likely to choose the studied or unstudied morph, but when they chose the studied morph, they were systematically more confident. The fact that we found a significant difference in participants' confidence in choosing between the studied and unstudied morphs illustrates that participants were not simply guessing when making these judgments. The confidence difference seems to

indicate that while the studied and unstudied morphs achieved equality in familiarity or evidence, they did so by potentially different routes or sources of information, which caused differences in confidence.

We needed to rule out the possibility that this confidence–recognition dissociation was due simply to just a few atypical faces. Figure 4 shows the scatterplot of the choosing rates for the studied and unstudied morphs plotted against the mean confidence when each was chosen. There are 16 points of each type because we graphed left and right presentations separately. We did so because while we did not see a left–right bias for all the morphs, we did want to look for this possibility in individual morph pairs. As can be seen in Figure 4, the probability of choosing a particular face was balanced across the graph, which led to the overall probability of choosing the studied morph equal to the 0.4994 value we observed, but studied morphs were consistently given higher confidence ratings when chosen. This occurred even for studied morphs that were chosen only rarely, and thus this effect occurred over the entire range of data, as shown by the linear regression lines fit to the studied and unstudied morph data. Thus, the confidence–recognition dissociation occurred at the level of individual items and was not merely an artifact of averaging or a technical issue with the experimental analyses.

Failure of SDT

The traditional extension of SDT to confidence ratings is to include additional decision criteria along the decision axis for each confidence level (Macmillan & Creelman, 1991). As targets become stronger and acquire more evidence, the distribution of target evidence values shifts to the right, which increases the hit rate for these strong targets and also moves more area into higher confidence regions along the evidence axis. Thus, we typically find that as the hit-rate increases for a condition, confidence increases as well, and this is entirely consistent with SDT.

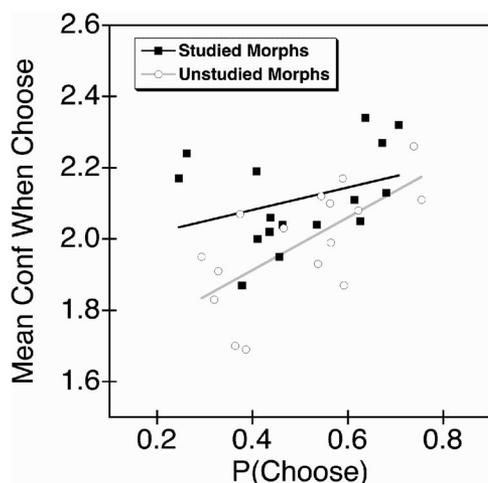


Figure 4. Choosing rate for studied and unstudied morphs plotted against the mean confidence when each is chosen. The lines represent best-fitting linear regression lines for the two types of morphs. The studied morphs have systematically higher confidence values than the unstudied morphs, even when the choosing rates are quite low (left side of the graph).

The evidence axis in two-alternative forced choice is a little different than in old/new recognition. In two-alternative forced choice, the evidence axis is the difference between the evidence for the right face and the evidence for the left face. The left panel of Figure 5 illustrates this representation. When the target item appears on the right side of the display, it typically has more evidence than the distractor item appearing on the left, and thus the difference in strength or evidence is greater than zero (black curve). If the target appears on the left, the distribution of evidence differences is shifted below zero (gray curve). Target items are associated with higher confidence in this representation because in order for the distractor item to be chosen on trials in which the target is on the right, the evidence difference would have to be negative. As can be seen in the left panel of Figure 5, there is relatively little area to the left of zero under the black curve and almost none in the high-confidence region on the left side. Thus, SDT tends to predict that confidence will be higher when participants are choosing targets than when they are choosing distractors, which is an entirely reasonable prediction.

As the targets become more discriminable from the distractors (say by increasing participants' study time), the evidence differences grow, and these two curves symmetrically shift away from zero in opposite directions. This produces more area under the curves in the higher confidence regions, and thus SDT can account for the typical finding that conditions that produce higher accuracy also produce higher confidence.

Now consider the condition in which the evidence for the studied and unstudied morphs is equated, as we saw in the data. This is where Gaussian SDT breaks down. If the participants do not have left–right biases (and we found none), a choosing rate of 0.5 for studied and unstudied morphs produces two symmetric distributions with the decision criterion at zero. Having accomplished this, however, the model cannot produce a confidence difference between the studied and unstudied morphs because the curves are right on top of each other. There is no way for distractor items to have less area in higher confidence regions than target items due to the symmetry of the model.

Targets and distractor distributions with unequal variances cannot save the model because in two-alternative forced choice, the familiarity of the distractor item is subtracted from the familiarity of the target item. This occurs for both targets presented on the right side and targets presented on the left side. The net result is two symmetric distributions that are also Gaussian and are right on top of each other in order to produce choosing rates that are at .5 for studied and unstudied morphs. In fact, the example graph in right panel of Figure 5 is consistent with target and distractors distributions that have unequal variances but identical means.

To summarize, the fact that we found equal choosing rates for studied and unstudied morphs but higher confidence when observers chose the studied morph represents a property test of the unequal variance signal detection model, and it fails this test.

To verify that this logic is correct, we conducted quantitative fits of the SDT model. The left panel of Figure 6 shows the receiver operating characteristic (ROC) curves for filler and morph faces, along with the best-fitting values from three different models. The axes on the ROC plots may look unfamiliar at first, but the way ROC curves are constructed in two-alternative forced choice is that one side is picked arbitrarily as a reference (in this case, the left side). In traditional yes/no recognition, ROC curves hold the

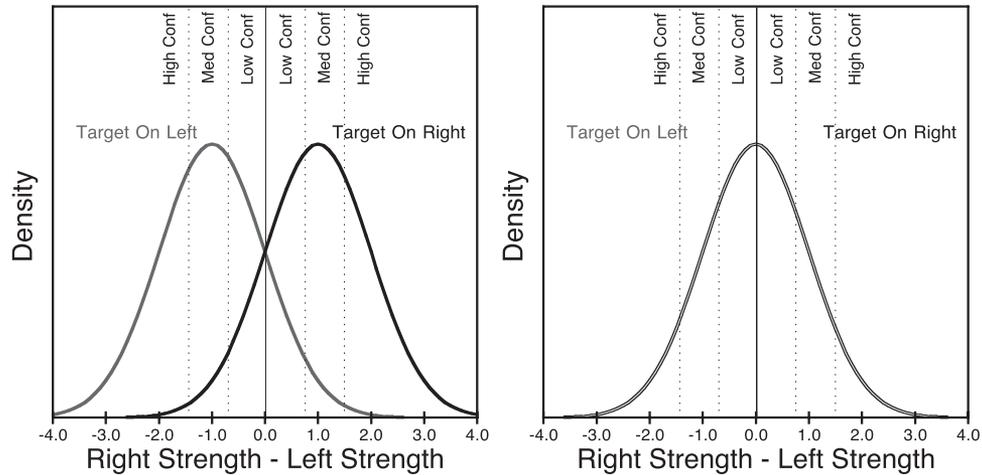


Figure 5. Representation of signal detection theory for two-alternative forced-choice experiments. Left panel: The evidence axis represents the difference in strength between the item presented on the right of the screen and the item on the left of the screen. If the target is presented on the right, it typically has more strength than the item on the left, and thus the values tend toward the positive end of the scale (back curve). The reverse is true for targets presented on the left side of the screen, which tend to produce negative values (gray curve). Confidence criteria are placed along this axis, and the area under the curve inside each region gives the percentage rating. Right panel: If targets and distractors have equal strengths (as is the case for our studied and unstudied morphs), they are represented in the model as two curves that overlap (double-lined curve). The symmetry of the model prevents it from assigning higher confidence values to targets than to distractors, thus making it unable to account for the confidence–recognition dissociation. See text for details.

response constant and vary the stimulus to make the vertical axis the probability of a participant saying “old” given a target and the horizontal axis the probability of a participant saying “old” given a distractor. In two-alternative forced choice, we hold one response constant (saying “left”) and vary whether the target is on the left to give the analog to hits or whether the target is on the right to give the analog of false alarms.

To fit the SDT model, we assumed that morphed and filler faces had different means and standard deviations for their evidence

distributions. This makes sense because the filler faces had higher choosing rates when they were targets than the morphs did. Because the evidence axis is the difference between the target on the left and the target on the right for two-alternative forced-choice SDT, it is redundant to fit separate variances for target and distractor stimuli. However, because morphs may have different properties than filler faces (perhaps they are more homogeneous as a group), it was reasonable to assume that the morphs should have their own variance for their distribution.

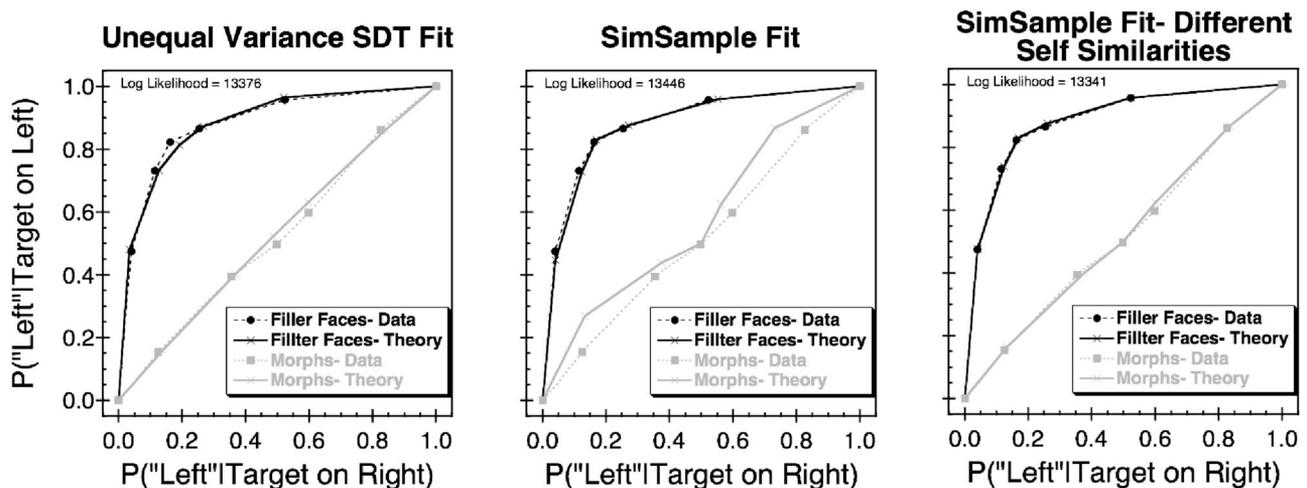


Figure 6. Receiver-operator-characteristic plots with model fits from the unequal variance signal detection theory (SDT) model (left panel) and the SimSample model without separate self-similarity values (center panel) and with separate self-similarities for morphs and filler faces (right panel).

The model has six free parameters to fit 20 data points, which were derived from two types of test trials (targets on the left and targets on the right), crossed with two types of items (morphs and filler faces) crossed with five unique levels of confidence (the sixth level is not free to vary). The parameters were the mean of the morph distributions, the mean of the filler distributions, the ratio of variances between the morphs and filler faces, and two confidence criteria. We assumed the confidence criteria were symmetric and therefore reflected them across zero. We also fitted a parameter that indicated the location of the decision criterion on the evidence difference axis, which, given the relative lack of left/right bias discussed earlier, was likely to be very close to zero. We estimated the parameters using Solver in Excel (Microsoft Corp., Redmond, WA), which maximized the log likelihood of the data given a particular set of parameters. The left panel of Figure 6 shows the estimates from the best-fitting parameters of the model, and Table 1 shows the relevant choosing rates and confidence values for morphs and filler faces that were calculated from the fitted data points.

The overall fit was not bad, but there are problems in specific areas. In the data, there was a systematic difference in confidence of about 0.11 between the studied and unstudied morphs (2.13 vs. 2.02). Unfortunately, the SDT fit could manage only a difference of 0.03 for these two conditions. In fact, the model accomplished this modest difference only by cheating: The choosing rate for studied morphs must be set at a value that is too high (0.518 vs. the 0.4994 seen in the data) to produce even this modest difference in confidence. This implies that the model is attempting to distribute error variance across the different measures, and the compromise fits neither the choosing rates nor the confidence values. When the mean of the morph distribution was set to zero to force the choosing rates to go to 0.5 for the studied and unstudied morphs, the model produced a difference between studied and unstudied morph confidence of exactly zero. These simulations verified the logic discussed previously that Gaussian SDT cannot account for the confidence–recognition dissociation seen with the studied and unstudied morphs.¹

Recollection and Familiarity

One possible explanation for the confidence–recognition dissociation seen with the morphs is that studied morphs might have been chosen primarily on the basis of recollection, while unstudied morphs were chosen primarily on the basis of familiarity. If recollective processes tend to be associated with higher confidence responses than responses made on the basis of familiarity, this could account for the differences. The remember/know judgment may not be process pure (Rotello, Macmillan, Reeder, & Wong, 2005), but it may still be revealing. We included a remember/know response in our testing procedure, which occurred after the participants made their combined choice/confidence response. We calculated the rate of responding “remember” for studied and unstudied morphs for each participant and found only a small, nonsignificant difference for the morphs (0.363 for studied morphs and 0.340 for unstudied morphs, with a 95% confidence interval of [−0.016, 0.067] on the difference that includes zero).

The finding of no significant difference in remember responses for studied versus unstudied morphs offers no direct support for a recollection/familiarity account. However, the suggestion that par-

ticipants make contact with individual items in memory is an attractive one, and in the following section, we propose an alternative model that is a single-process model yet includes elements of what might be considered recollective processes.

An Alternative Model: SimSample

The confidence–recognition dissociation illustrates the role that similarity plays in both recognition and confidence. The SimSample model was proposed previously by Busey and Tunnicliff (1999) explicitly to model the effects of similarity in recognition experiments with faces. In this model, both recognition and confidence are related to the similarity between the test face and sampled face. If the two are more similar, the observer is more likely to say “old” and will be more confident when doing so. The sampling process is related to the similarity between the test face and other faces such that similar faces are more likely to be sampled—thus, the name SimSample. Although the SimSample model applied the sampling process to faces, many of the model assumptions are based on the SAM model (Gillund & Shiffrin, 1984) and the GCM (Nosofsky, 1986). Both of these models have been extremely successful in accounting for various phenomena in their respective areas, and thus many of the assumptions that underlie SimSample have been validated in previous work.

SimSample Accounts of the Confidence–Recognition Dissociation

SimSample works as follows. At test, the test face (or each face sequentially in forced choice) is compared against the contents of memory through a sampling process. This sampling process happens only once, and it recovers exactly one face from memory to compare against the test face. The evidence that the test face was studied is related to the similarity between the test face and the sampled face, and confidence is also based only on the similarity between the test face and the sampled face. In a forced-choice design, each test face is treated separately in the sampling process, and the evidence is combined into a single decision. It is important to note that the sampling process is more likely to recover faces that are similar to the test face, so dissimilar faces are sampled only occasionally. Of course, when dissimilar faces are sampled, they produce low confidence.

The SimSample model might account for the confidence–recognition dissociation according to the following logic, as illustrated in Figure 7. The studied morph can sample itself in memory, which would lead to strong evidence that it was studied. However, if it does not sample itself, it must sample something else like a filler face, which will produce much weaker evidence. The unstudied morph has two somewhat weaker matches and therefore has two opportunities to avoid sampling a very dissimilar filler face although the evidence when a parent face is sampled is weaker than when the studied morph samples itself. If things are balanced correctly, two nearby matches for the unstudied morph (its parents) could be equivalent to a single strong match for the studied morph (itself).

¹ We tried a several non-Gaussian distributions including the highly asymmetric exponential but could not find a combination that had both a reasonable quantitative fit and a set of reasonable parameter values.

Table 1
Choosing Rates and Mean Confidence When Choosing Each Stimulus Type for the Data and Three Candidate Models

Stimulus type	Data		Unequal-variance SDT		SimSample		SimSample: Separate morph self-similarities	
	Choosing rates	Mean confidence when choosing	Choosing rates	Mean confidence when choosing	Choosing rates	Mean confidence when choosing	Choosing rates	Mean confidence when choosing
Studied morphs	0.499	2.13	0.518	2.09	0.500	2.42	0.500	2.13
Unstudied morphs	0.501	2.02	0.482	2.05	0.500	2.02	0.500	2.02
<i>Difference</i>	<i>-0.001</i>	<i>0.11</i>	<i>0.036</i>	<i>0.03</i>	<i>-0.001</i>	<i>0.40</i>	<i>-0.001</i>	<i>0.11</i>
Filler targets	0.830	2.46	0.811	2.50	0.830	2.42	0.830	2.46
Filler distractors	0.170	1.97	0.189	1.86	0.170	1.97	0.170	1.97

Note. The italicized row contains the differences between the studied and unstudied morphs, which may not be exact due to rounding errors in the table. SDT = signal detection theory.

However, the sampling model assumes that confidence is based only on the similarity between the sampled face and the test face. The unstudied morph benefits from two nearby parents, but only one can be sampled on each trial. Because a parent face is less similar to the unstudied morph than the studied morph is to itself, the studied morph will receive higher confidence ratings when it is chosen.

Note that other models such as a summed similarity model would predict that confidence is determined by the summed similarity to all faces in memory, rather than just to the sampled face. This would lead to a close correspondence between confidence and recognition rates and could not account for the confidence–recognition dissociation.

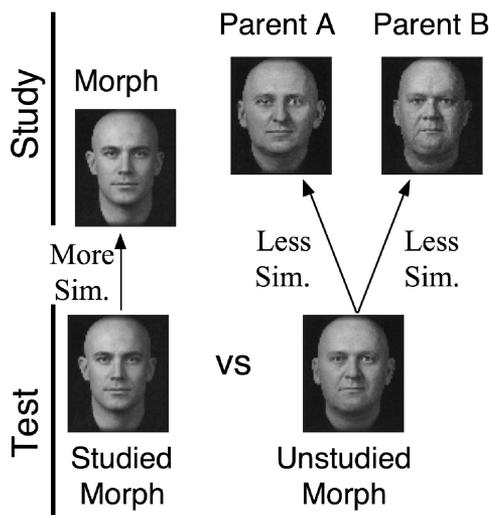


Figure 7. Sampling model explanation of the confidence–recognition dissociation. The studied morph is more similar to one item in memory (itself), while the unstudied morph is less similar to two items in memory (its parents). This makes choosing rates equivalent because two nearby items come close to equaling the evidence from one strong match. However, only one face is ultimately sampled for each test face, and confidence is determined only by the similarity between the chosen test face and the sampled face, making confidence lower in the case of the unstudied morph when it is chosen.

Quantitative Fits of the SimSample Model

To verify that the logic described actually allows the SimSample model to account for the confidence–recognition dissociation, we tested the model using quantitative fitting procedures. We chose to model the similarity between items as free parameters. In related work, authors have chosen to derive similarity from distances in some psychological face space (Busey, 1998; Nosofsky, 1986; O’Toole, Abdi, Deffenbacher, & Valentin, 1993) using a nonlinearity such as an exponential (e). In this analysis, we simply modeled distance directly as a free parameter and computed similarity as follows:

$$\eta_{i,j} = e^{cd_{i,j}} \tag{1}$$

where $d_{i,j}$ is the distance between faces i and j , and $\eta_{i,j}$ is the similarity between faces i and j . The parameter c is a scaling parameter. As distance increases, the similarity value drops exponentially toward zero.

Our design structure required us to model the distance of an item to itself (and to start with, we fixed this at zero), the distance of unstudied morphs to their parents, and the average distance of all morphs to the filler faces. We also chose to limit model testing to the combined data across all morphs rather than fit individual pairs of morphs because the overall data provided the dissociation that disconfirmed the SDT approach and because combining across items provided the most power to discriminate between candidate models.

Sampling From Memory to Provide Choosing Rates

SimSample assumes that the test item is used to sample memory, and one item is ultimately sampled and compared with the test item (previous modeling attempts used multiple samples without improvement in the fits; see Busey & Tunnicliff, 1999). The sampling is not done uniformly but instead on the basis of similarity, such that similar items are more likely to be sampled. The probability that the observer samples face k in memory, given that face i was presented at test is computed as follows:

$$P(\text{sample } k | i \text{ presented}) = \frac{\eta_{i,k}}{\sum_{j \in \text{All Faces In Memory}} \eta_{i,j}} \tag{2}$$

where $\eta_{i,k}$ is the similarity between faces i and k . This is simply the Luce choice rule.² SimSample was originally designed to handle old/new data and then was extended to forced-choice data (Busey, 2001; Busey & Tunnicliff, 1999).

In the forced-choice version of the model, we assumed that the evidence computed between a sampled face and the test face was a nonlinear (squashing) function of the similarity. The evidence was computed as follows:

$$\Theta(\eta_{i,k}) = \int_{-\infty}^{\eta_{i,k}} \frac{e^{-(x-criterion)^2/2\text{critSD}^2}}{\sqrt{2\pi\text{critSD}^2}} dx \quad (3)$$

which implied that if $h_{i,k}$ equaled the criterion, evidence in favor of this test face when face k was sampled was 0.5 and reduced as the similarity became less. This gives two parameters related to the computation of evidence: the criterion and the standard deviation of the criterion (critSD). Note that the specific form of this non-linearity was less important; a logistic function would have produced very similar results. In the old/new variant of SimSample, the criterion and critSD had the interpretation of adding decisional noise to the decision.

For forced-choice paradigms, we assumed that the participant computed the evidence in favor of each test and then chose the face with the highest evidence of being on the list. However, a long history of research has documented the fact that this process is not done in an all-or-none fashion and instead requires a softening function to allow for probability matching (Maddox & Ashby, 1993; Nosofsky & Zaki, 1998). Thus for two faces a and b , with evidence A and B , respectively, computed from Equation 3, we computed the choosing rate for face a as follows:

$$P(\text{Choose } a \text{ and } b) = \frac{e^{\zeta A}}{e^{\zeta A} + e^{\zeta B}} \quad (4)$$

This required a third free parameter, ζ , which was the probability matching parameter.

Confidence Ratings

Once an item is chosen, the model must also generate a confidence rating, which up until now has not been formalized in the literature. A variety of possibilities exist, from using the overall density of the faces around the test face (Cottrell, Dailey & Busey, 1998), the similarity of the two faces to each other (Chandler, 1994; Clark, 2003; Tulving, 1981), and the differences between the two evidence values (Van Zandt, 2000). Each of these factors may play some role. However, for reasons of parsimony, we chose to rely on the same evidence that went into the choice decision: the similarity of the sampled face to the test face.

There were three possible confidence ratings that participants could use for the face they chose as having been on the study list. To model these ratings, we assumed that there were two confidence criteria that partitioned the similarity scale into three intervals. We also assumed decisional noise in this process. The decisional noise creates a normally distributed set of values such that the similarity between the sampled face and the chosen face take on different values across trials, items, and participants. As we did with the probability of saying “old” (which ultimately affects the

choosing rates), we computed the probability of giving a confidence rating response of 1 as the area under this curve that falls to one side of a decision criterion as follows:

$$C_1(\eta_{i,k}) = \int_{-\infty}^{\eta_{i,k}} \frac{e^{-(x-ConfCrit_1)^2/2ConfSD^2}}{\sqrt{2\pi\text{confSD}^2}} dx \quad (5)$$

Put another way, this is the area under a normally distributed curve centered at the similarity between the two faces, to the left of some criteria (meaning items that are not very similar to the test item tended to receive low confidence values fairly often). The probability of receiving a high confidence rating of 3 was computed similarly:

$$C_3(\eta_{i,k}) = 1 - \int_{-\infty}^{\eta_{i,k}} \frac{e^{-(x-ConfCrit_3)^2/2ConfSD^2}}{\sqrt{2\pi\text{confSD}^2}} dx, \quad (6)$$

which is simply the area to the right of some criterion under the normal curve centered on the similarity between the sampled face and the test face. Sampled faces that are most similar to the test face tend to receive high ratings because the curve shifts, pushing more area into the high-confidence region.

The probability of giving a response of 2 can easily be computed from the other two:

$$C_2(\eta_{i,k}) = 1 - C_3(\eta_{i,k}) - C_1(\eta_{i,k}) \quad (7)$$

because the participant must give one of these three responses.

The previously described parameters are independent of the stimulus type (filler or morph) and are therefore assumed to be constant regardless of whether an item was studied or not, or whether it was a filler face or a morph. To obtain the values of similarity, we estimated the distances in psychological space between different items. The distance between an item and itself in memory, $d_{i,j}$, was assumed in this version of the model to be set to zero (although an extension will amend this for some items). The distances between an unstudied morph and its parents, $d_{\text{morph, parents}}$, and the average distance between one face and all others, $d_{i, \text{others}}$, were the final two free parameters in the model.

To summarize, there are eight free parameters, two more than the signal detection model. However, this model has additional assumptions, which allow it to account for elements of the data that are driven by the similarity of individual items, and it needs these additional parameters to handle this additional complexity.

This version of the SimSample model produced mixed results, as shown by the middle panel of Figure 6 and the summary statistics shown in Table 1. On the positive side, it nicely captured the difference in confidence between the studied and unstudied morphs and does so without forcing the choosing rates for the studied morphs away from 0.5, as we saw with the SDT fit. On the negative side, the amount of difference between the two conditions is almost four times as great as what we observed in the data. The

² The sampling rule proposed in Gillund and Shiffrin’s (1984) SAM model also uses strengths to compute the Luce choice ratio.

middle panel of Figure 6 also illustrates that the model's fit to several points in the morph data was poor, and the overall fit was worse than the signal detection account. Perhaps ironically, this failure does illustrate one important point: Just because one model has more free parameters than another does not guarantee a better fit. Table 2 provides the fit statistics for all models under comparison.

Given these results, the SimSample model could be viewed as a qualitative success (it does capture the essential elements of the confidence–recognition dissociation) and a quantitative failure. There are various reasons that the model might fail and several possible solutions. The problem that most readily comes to mind is that the morphs may differ in some fundamental way from the filler faces, and in an extension we explore how morphs might be less similar to their images in memory than other faces are to their own images.

Morph Self-Similarity

In the original version of SimSample, we assumed that morphs and filler faces share similar characteristics in terms of self-similarity. However, morphs may systematically lie in denser regions (Busey, 1998), may lack distinctive features that tend to get blended away during morphing, and may be more confusable with other stimuli. Other authors have dealt with this by assuming that different types of faces can have different self-similarities (Knapp, Nosofsky, & Busey, 2006; Nosofsky & Zaki, 2003) or different kernel densities (Dailey et al., 1998). Nosofsky and Zaki (2003) made color chips more distinctive by adding a symbol to several chips during the study period and found that they could account for the enhanced memory effect by increasing the self-similarity parameter for these stimuli. Knapp et al. (2006) generalized this finding to faces and found that distinctive faces required greater self-similarity. A distinctive face may strongly match itself in memory and be more self-similar, while a more typical face like a morph may be subject to more noise and therefore be further away in psycho-

logical space. To model this in the present experiment, we assumed that the distance between a studied morph and itself in memory was not zero, but a free parameter, $d_{\text{morph, morph}}$, that is greater than zero (i.e., more distant and therefore less similar) than those for distinctive faces. This added a ninth free parameter to the model but does not otherwise change the model structure. Because we had 20 data points, the model was nowhere near saturated, even with this additional parameter.

As this parameter moved away from zero, the similarity between the studied morph and itself in memory dropped below 1.0, which would affect both the choosing rates and the confidence values. How these are affected was complicated because other parameters also changed as a result of the addition of this parameter.

This model did an excellent job of accounting for the data at both the qualitative and quantitative levels. The right panel of Figure 6 illustrates the fit, which was better than the SDT fit and much better than the original SimSample model fit. The rightmost columns of Table 1 illustrate that not only did the model accurately capture the equality of the choosing rates for studied and unstudied morphs, it also accurately captured the confidence values when each was chosen. In fact, the values obtained for the choosing rates and the confidence when each was chosen were almost identical to the data values. This shows that this version of SimSample can quantitatively account for the confidence–recognition dissociation that proved so problematic for SDT.

The SimSample model with the different self-similarities extension has nine free parameters compared with the six found in SDT. One way to compare these models is to use the Akaike information criterion and Bayesian information criterion, which both aim to penalize the model with more parameters. This extended version of SimSample had better values of both Akaike information criterion and Bayesian information criterion than the signal detection fit, as illustrated by Table 2. Thus, the extra parameters used in the extended SimSample model seem justified, along with the particular architecture expressed in the model.

Table 2
Best-Fitting Parameters, Log-Likelihood Values, and Akaike Information and Bayesian Information Criteria for the Three Models

Parameter	Unequal-variance SDT	Parameter	SimSample	SimSample–two self-similarities
ConfCrit1	–1.043	Distance to parents	0.397	0.567
ConfCrit2	–0.290	Distance to fillers	0.441	0.631
ConfCrit3	–0.011	Probability matching	7.536	11.84
Morph mean	–0.045	Crit	0.635	0.503
Filler mean	–0.987	CritSD	0.020	0.036
Filler variance	1.121	ConfCrit ₁	0.477	0.324
		ConfCrit ₂	0.834	0.770
		ConfCritSD	0.272	0.340
		Morph distance to self	—	0.383
Log likelihood	13376.3		13446.8	13341.0
No. of parameters	6		8	9
AIC	26770.5		26909.5	26700.0
BIC	26764.6		26917.5	26708.9

Note. For all three criteria, smaller numbers are better for each row, and the SimSample model with two self-similarities is a better fit by all three criteria. ConfCrit = confidence criterion; SDT = signal detection theory; AIC = Akaike information criterion; BIC = Bayesian information criterion.

General Discussion

We began this article with the suggestion that two types of stimuli could be equated on one dimension and show differences on another dimension. We argued that this result could be diagnostic of the manner in which the two stimuli reached equivalence. Through careful manipulation of similarity of parent faces to unstudied morphs, we were able to equate the choosing rates for studied and unstudied morphs in a forced-choice test and then show that observers were systematically more confident when choosing the studied morph. This result proved incompatible with even an unequal variance Gaussian SDT account. However, we were able to develop an extended version of a sampling model, SimSample, which captured both the qualitative and quantitative aspects of the confidence–recognition dissociation. These results are consistent with a model in which individual traces in memory are accessed during the recognition process. This contact provides a strong match to memory in the case of the studied morph and also higher confidence. The unstudied morph is chosen on the basis of a weaker match to memory that tends to lower confidence, but unstudied morphs have two nearby weak matches, which boost the choosing rate to around that of the studied morph. Thus two lesser matches are equivalent to one stronger match when it comes to choosing rates due to the nature of the sampling process, but the sampling statistics do not affect confidence, which is based solely on the similarity between the sampled item and the test face.

This access to the strength of the match to individual items suggests that observers are using a recall-like process, as opposed to one that computes only a global memory match or overall familiarity measure. This finding is consistent with that of Yonelinas, Kroll, Dobbins, and Soltani (1999), who also argued for a recall process for faces. The current results lead to a similar conclusion through very different methods and extend the result to suggest that the recall process can make a contribution to the recognition of even very typical faces such as the morphs. Thus, it is not limited to recalling very distinctive items with high confidence. We put forth two lines of evidence in favor of this conclusion: both the failure of the property test for SDT and the success of the sampling model in accounting for the dissociation between confidence and recognition. The success of the sampling model should not be viewed in isolation (see Roberts and Pashler, 2000, for a discussion of interpreting model fits) but in conjunction with the property test results.

A growing body of literature has dissociated familiarity and recollection in visual memory tasks, which was put forth initially by Jacoby (1991) and extended to feature and conjunction errors by Jones and Jacoby (2001). Curran and Cleary (2003) used event-related potentials (ERPs) to dissociate the two processes, while in a related series of studies, Curran and Dien (2003) and Curran (2004) looked at the role of attention and modality-specific processing in recollection and familiarity using ERP procedures. Similar work was done with verbal materials by Rugg et al. (1998). Yonelinas (1994, 1997, 2001) had a series of articles on this topic and an excellent review (Yonelinas, 2002). The relation between the sampling model's access to individual items and the recollection/familiarity debate must be tempered by the finding that observers did not give systematically more "remember" responses to studied morphs than to unstudied morphs when they chose each. This suggests that, at least as operationalized by our recollection/

familiarity instructions, observers did not have conscious awareness of the use of different strategies for studied and unstudied morphs. These null effects are entirely consistent with the single-process account of SimSample.

Implications for Models of Metamemory

The sources of information that contribute to metacognitive judgments such as confidence ratings can be roughly divided into what can be thought of as heuristics and those that derive from memorial sources. Heuristics include factors such as "Brighter stimuli always produce better performance" and "I would have remembered this had I seen it, and so I'm confident I didn't see it." These factors are often viewed as outside the domain of extensions of memory models that account for confidence. The second source of information comes from extensions to memory models that originally just accounted for memory performance (e.g. SAM, Raaijmakers & Shiffrin, 1980, 1981; Minerva 2, Hintzman, 1986; CHARM, Eich, 1982, 1985; REM, Shiffrin & Steyvers, 1997; TODAM, Murdock, 1982). Norman and O'Reilly (2003) have used elements of a neural network model to suggest how the contributions of hippocampal and neocortical structures might account for effects similar to those reported here, including memory sensitivity and specificity. These approaches indicate the need to consider the relation between hypothesized representations and the processes that interact on them when one is accounting for memory results and extensions to metacognitive judgments, before resorting to extramemorial sources such as heuristics.

References

- Arndt, J., & Reder, L. M. (2002). Word frequency and receiver-operating characteristic curves in recognition memory: Evidence for a dual-process interpretation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 830–842.
- Backus, B. T., & Nolt, M. J. (2001). Analysis of stereoscopic metamers [Abstract 179]. *Journal of Vision*, 1, 179a. doi:10.1167/1.3.179
- Banks, W. P. (2000). Recognition and source memory as multivariate decision processes. *Psychological Science*, 11, 267–273.
- Berry, C. J., Shanks, D. R., & Henson, R. N. A. (2008). A unitary signal-detection model of implicit and explicit memory. *Trends in Cognitive Sciences*, 12, 367–373.
- Busey, T. (2001). Formal models of familiarity and memorability in face recognition. In M. Wenger & J. Townsend (Eds.), *Computational, geometric, and process perspectives on facial cognition*. Hillsdale, NJ: Erlbaum.
- Busey, T. A. (1998). Physical and psychological representations of faces: Evidence from morphing. *Psychological Science*, 9, 476–483.
- Busey, T. A., Tunnicliff, J., Loftus, G. R., & Loftus, E. (2000). Accounts of the confidence–accuracy relation in recognition memory. *Psychonomic Bulletin & Review*, 7, 26–48.
- Busey, T. A., & Tunnicliff, J. L. (1999). Accounts of blending, distinctiveness, and typicality in the false recognition of faces. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 1210–1235.
- Cary, M., & Reder, L. M. (2003). A dual-process account of the list length and strength-based mirror effects in recognition. *Journal of Memory and Language*, 49, 231–248.
- Chandler, C. C. (1994). Studying related pictures can reduce accuracy, but increase confidence, in a modified recognition test. *Memory and Cognition*, 3, 273–280.

- Clark, S. E. (2003). A memory and decision model for eyewitness identification. *Applied Cognitive Psychology, 17*, 629–654.
- Curran, T. (2004). Effects of attention and confidence on the hypothesized ERP correlates of recollection and familiarity. *Neuropsychologia, 42*, 1088–1106.
- Curran, T., & Dien, J. (2003). Differentiating amodal familiarity from modality-specific memory processes: An ERP study. *Psychophysiology, 40*, 979–988.
- Curran & Cleary. (2003). Using ERPs to dissociate recollection from familiarity in picture recognition. *Cognitive Brain Research, 15*, 191–205.
- Dailey, M., Cottrell, G., & Busey, T. (1999). Facial memory is kernel density estimation (almost). In Michael J. Kearns, Sara A. Solla, & David A. Cohn (Eds.), *Advances in Neural Information Processing Systems (NIPS) 11: NIPS Conference, Denver, CO, November 30–December 5, 1998*. Cambridge, MA: MIT Press.
- DeCarlo, L. T. (2007). The mirror effect and mixture signal detection theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 18–33.
- Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap* (Monographs on Statistics and Applied Probability, Vol. 57). New York: Chapman & Hall/CRC.
- Eich, J. M. (1982). A composite holographic associative recall model. *Psychological Review, 89*, 627–661.
- Eich, J. M. (1985). Levels of processing, encoding specificity, elaboration, and CHARM. *Psychological Review, 92*, 1–38.
- Franks, J. J., & Bransford, J. D. (1971). Abstraction of visual patterns. *Journal of Experimental Psychology, 90*, 65–74.
- Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91*, 1–67.
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- Hart, J. T. (1967). Memory and the memory-monitoring process. *Journal of Verbal Learning and Verbal Behavior, 6*, 685–691.
- Hintzman, D. L. (1986). “Schema abstraction” in a multiple trace memory model. *Psychological Review, 93*, 411–428.
- Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95*, 528–551.
- Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30*, 513–541.
- Jones, T. C., & Jacoby, L. L. (2001). Feature and conjunction errors in recognition memory: Evidence for dual-process theory. *Journal of Memory and Language, 45*, 82–102.
- Kayser, A. (1985). *Heads*. New York: Abbeville Press.
- Knapp, B. R., Nosofsky, R. M., & Busey, T. A. (2006). Recognizing distinctive faces: A hybrid–similarity exemplar model account. *Memory & Cognition, 34*, 877–889.
- Koriat, A. (1993). How do we know what we know? The accessibility model of the feeling of knowing. *Psychological Review, 100*, 609–639.
- Koriat, A. (1997). Monitoring one’s own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126*, 349–370.
- Loftus, G. L., & Ruthruff, E. (1994). A theory of visual information acquisition and visual memory with special application to intensity-duration tradeoffs. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 33–49.
- Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user’s guide*. New York: Cambridge University Press.
- Maddox, W. T., & Ashby, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Perception & Psychophysics, 53*, 49–70.
- Masson, M. E. J., & Loftus, G. R. (2003). Using confidence intervals for graphically based data interpretation. *Canadian Journal of Experimental Psychology, 57*, 203–220.
- Meissner, C. A., Tredoux, C. G., Parker, J. F., & MacLin, O. H. (2005). Eyewitness decisions in simultaneous and sequential lineups: A dual-process signal detection theory analysis. *Memory & Cognition, 33*, 783–792.
- Mickes, L., Wixted, J. T., & Wais, P. (2007). A direct test of the unequal variance signal detection model of recognition memory. *Psychonomic Bulletin & Review, 14*, 858–865.
- Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review, 80*, 609–625.
- Neumann, P. G. (1977). Visual prototype formation with discontinuous representation of dimensions of variability. *Memory & Cognition, 5*, 187–197.
- Norman, K. A., & O’Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: A complementary learning systems approach. *Psychological Review, 110*, 611–646.
- Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115*, 39–57.
- Nosofsky, R. M., & Zaki, S. R. (1998). Dissociations between categorization and recognition in amnesic and normal individuals: An exemplar-based interpretation. *Psychological Science, 9*, 247–255.
- Nosofsky, R. M., & Zaki, S. R. (2003). A hybrid-similarity exemplar model for predicting distinctiveness effects in perceptual old-new recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 1194–1209.
- O’Toole, A., Abdi, H., Deffenbacher, K., & Valentin, D. (1993). Low-dimensional representations of faces in higher dimensions of the face space. *Journal of the Optical Society of America, A: Optics image science, and vision, 10*, 405–411.
- Parks, C. M., & Yonelinas, A. P. (2007). Moving beyond pure signal detection models: Comment on Wixted (2007). *Psychological Review, 114*, 188–202.
- Pelli, D. G., & Zhang, L. (1991). Accurate control of contrast on micro-computer displays. *Vision Research, 31*, 1337–1360.
- Raaijmakers, J. G. W., & Shiffrin, R. M. (1980). SAM: A theory of probabilistic search of associative memory. *Psychology of Learning and Motivation, 14*, 207–261.
- Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88*, 93–134.
- Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review, 107*, 358–367.
- Rotello, C. M., Macmillan, N. A., Hicks, J. L., & Hautus, M. J. (2006). Interpreting the effects of response bias on remember–know judgments using signal-detection and threshold models. *Memory & Cognition, 34*, 1598–1614.
- Rotello, C. M., Macmillan, N. A., & Reeder, J. A. (2004). Sum-difference theory of remembering and knowing: A two-dimensional signal-detection model. *Psychological Review, 111*, 588–616.
- Rotello, C. M., Macmillan, N. A., Reeder, J. A., & Wong, M. (2005). The remember response: Subject to bias, graded, and not a process-pure indicator of recollection. *Psychonomic Bulletin & Review, 12*, 865–873.
- Rugg, M. D., Mark, R. E., Walla, P., Schloerscheidt, A. M., Birch, C. S., & Allan, K. (1998, April 9). Dissociation of the neural correlates of implicit and explicit memory. *Nature, 392*, 595–598.
- Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM-retrieving effectively from memory. *Psychonomic Bulletin and Review, 4*, 145–166.
- Solso, R. L., & McCarthy, J. E. (1981). Prototype formation of faces: A case of pseudo-memory. *British Journal of Psychology, 72*, 499–503.
- Tulving, E. (1981). Similarity relations in recognition. *Journal of Verbal Learning and Verbal Behavior, 20*, 479–496.
- Van Zandt, T. (2000). ROC curves and confidence judgments in recogni-

- tion memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 582–600.
- Wixted, J. T. (1992). Subjective memorability and the mirror effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 681–690.
- Wixted, J. T. (2007a). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review*, 114, 152–176.
- Wixted, J. T. (2007b). Spotlighting the probative findings: A reply to Parks and Yonelinas. *Psychological Review*, 114, 203–209.
- Wixted, J. T., & Stretch, V. (2004). In defense of the signal-detection interpretation of remember/know judgments. *Psychonomic Bulletin & Review*, 11, 616–641.
- Wyszecki, G. (1958). Evaluation of metamer colors. *Journal of the Optical Society of America*, 48, 451–452.
- Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 1341–1354.
- Yonelinas, A. P. (1997). Recognition memory ROCs for item and associative information: The contribution of recollection and familiarity. *Memory and Cognition*, 25, 747–763.
- Yonelinas, A. P. (2001). Consciousness, control, and confidence: The 3 Cs of recognition memory. *Journal of Experimental Psychology: General*, 130, 361–379.
- Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, 46, 441–517.
- Yonelinas, A. P., Kroll, N. E. A., Dobbins, I. G., & Soltani, M. (1999). Recognition memory of faces: When familiarity supports associative recognition judgments. *Psychonomic Bulletin & Review*, 6, 654–661.
- Zaki, S. R., & Nosofsky, R. M. (2001). Exemplar accounts of blending and distinctiveness effects in perceptual old-new recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 1022–1041.

Received November 7, 2007

Revision received March 11, 2009

Accepted April 10, 2009 ■

Call for Nominations

The Publications and Communications (P&C) Board of the American Psychological Association has opened nominations for the editorships of **Experimental and Clinical Psychopharmacology**, **Journal of Abnormal Psychology**, **Journal of Comparative Psychology**, **Journal of Counseling Psychology**, **Journal of Experimental Psychology: Human Perception and Performance**, **Journal of Personality and Social Psychology: Attitudes and Social Cognition**, **PsycCRITIQUES**, and **Rehabilitation Psychology** for the years 2012–2017. Nancy K. Mello, PhD, David Watson, PhD, Gordon M. Burghardt, PhD, Brent S. Mallinckrodt, PhD, Glyn W. Humphreys, PhD, Charles M. Judd, PhD, Danny Wedding, PhD, and Timothy R. Elliott, PhD, respectively, are the incumbent editors.

Candidates should be members of APA and should be available to start receiving manuscripts in early 2011 to prepare for issues published in 2012. Please note that the P&C Board encourages participation by members of underrepresented groups in the publication process and would particularly welcome such nominees. Self-nominations are also encouraged.

Search chairs have been appointed as follows:

- **Experimental and Clinical Psychopharmacology**, William Howell, PhD
- **Journal of Abnormal Psychology**, Norman Abeles, PhD
- **Journal of Comparative Psychology**, John Disterhoft, PhD
- **Journal of Counseling Psychology**, Neil Schmitt, PhD
- **Journal of Experimental Psychology: Human Perception and Performance**, Leah Light, PhD
- **Journal of Personality and Social Psychology: Attitudes and Social Cognition**, Jennifer Crocker, PhD
- **PsycCRITIQUES**, Valerie Reyna, PhD
- **Rehabilitation Psychology**, Bob Frank, PhD

Candidates should be nominated by accessing APA's EditorQuest site on the Web. Using your Web browser, go to <http://editorquest.apa.org>. On the Home menu on the left, find "Guests." Next, click on the link "Submit a Nomination," enter your nominee's information, and click "Submit."

Prepared statements of one page or less in support of a nominee can also be submitted by e-mail to Emnet Tesfaye, P&C Board Search Liaison, at emnet@apa.org.

Deadline for accepting nominations is January 10, 2010, when reviews will begin.